19970314 074

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE

3. REPORT TYPE AND DATES COVERED
FINAL REPORT 01 Sep 92 - 31 Aug 96

4. TITLE AND SUBTITLE
Photonic Technology Development for Densely-Interconnected Neural Networks

5. FUNDING NUMBERS
62712E
7013/50

6. AUTHOR(S)
Professor B. Keith Jenkins

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
School of Engineering
Univ of Southern California
Los Angeles, CA 90089-2564

8. PERFORMING ORGANIZATION REPORT NUMBER
AFOSR-TR-~~96~~ 97
0142

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NE
110 Duncan Avenue Suite B115
Bolling AFB DC 20332-8050

F49620-92-J-0472

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Progress on the grant, "Photonic Technology Development for Densely Interconnected Neural Networks", is described. Substantial progress has been made in the following areas. Artificial neural-network models for implementation using photonics have been developed and analyzed. Radial basis function neural networks for analog nonparametric density function estimation and pattern recognition, and multilayer backward error propagation neural networks, both for implementation on photonic hardware, have been characterized. Emphasis has been on implementations based on incoherent/coherent double angular multiplexed volume hologram interconnected architectures using photorefractive materials, 2-D source arrays, and optoelectronic spatial light modulators, are emphasized. Our work on optoelectronic smart-pixel array spatial light modulators is also summarized. Silicon chips for arrays of neuron unit processing with optical inputs have been fabricated, as have GaAs-based chips incorporating inverted Fabry-Perot cavity strained-layer multiple-quantum-well modulators. Experimental results from are given. Results of flip-chip bonding using a newly developed Velcro-like indium bump bond process are given as well. Also included in this report are personnel supported, lists of publications and presentations, transitions to industry, and patent disclosures.

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	

Photonic Technology Development for Densely Interconnected Neural Networks

DARPA/AFOSR Grant No. F49620-92-J-0472

Combined

Final Technical Report, 9/1/92 – 8/31/96
Quarterly Technical Reports, 9/1/95 – 8/31/96

B. Keith Jenkins, P.I.
Anupam Madhukar, co-P.I.
Armand R. Tanguay, Jr., co-P.I.

Departments of Electrical Engineering and Materials Science
Center for Photonic Technology,
Center for Neural Engineering, and Signal and Image Processing Institute
University of Southern California
Los Angeles, California 90089-2564

Submitted 27 February 1997

1. Table of Contents

2. Executive Summary

2.1 Technical approach, accomplishments, and conclusions

A research and development effort on a photonic technology for the implementation of densely interconnected artificial neural networks has been undertaken. The purpose of the effort has been to develop the capability for implementation of flexible, modular, large-scale, highly parallel, nonlinear-analog neural networks, using a combination of optics and electronics. Our effort has spanned algorithms, neural-network models, photonic architectures, photonic components, and materials.

The photonic system we have developed and studied incorporates the following set of components and features. First, a 2-D array of individually coherent but mutually incoherent sources illuminates the photonic system. The key features of this array that are exploited in the system are the spatial distribution of sources and their mutual incoherence. Second, a set of optoelectronic smart-pixel spatial-light modulators (SLM's) are incorporated and are all based on the same flexible technology: a silicon chip that incorporates detection and nonlinear-analog electronic signal-processing functions, and a gallium-arsenide-based chip that incorporates optical modulation. These optoelectronic smart-pixel SLM's can implement a parallel array of neuron units, a parallel array of error-term-calculation units, and a parallel array of training-term generators, each having (where desired) optical input and optical output. Third, an optical interconnection system is based on a volume hologram for storage of interconnection patterns and weights, implemented in a real-time photorefractive material for adaptive neural-network processing, or in a fixed thick (or stratified) material for nonadaptive (*e.g.*, pre-trained) neural networks. The double angularly multiplexed geometry provides for a number of advantages, including reduced crosstalk during recording and readout, and the capability of optically copying the entire interconnection pattern into another volume hologram medium.

At the algorithm and system-architecture levels, we have developed neural-network models for implementation on a class of photonic system architectures. The models incorporate analog, nonnegative signals and adaptation or training capability. Applications to single and multilayer networks, including backward-error-propagation multilayer networks, and radial-basis-function-like networks for pattern recognition and density-function estimation, have been studied. Learning and computing behavior has been verified by theory and by simulation. In conjunction with partial support from the related AASERT grant (No. F49620-93-1-0445), we have modeled a set of salient characteristics of optical photorefractive-based implementations of multilayer backward-error-propagation neural networks (including such parameters as photorefractive

grating decay, spatial light modulator gain and output modulator contrast ratio, number representation and physical architecture, and detector limitations), and have derived a set of conditions for convergence of the learning process. This has also been verified by simulation.

At the device level, we have designed, fabricated, and characterized numerous silicon chips for neuron-unit input, signal processing, and output functions. Two-dimensional smart-pixel arrays of nonlinear-analog neuron-unit signal-processing circuits, with dual-channel optical inputs and dual-channel modulator-driver outputs, have been fabricated. Pixel sizes are 100 μm × 100 μm, their operation is extremely uniform over the array, and bandwidths are > 4 MHz for large signals and > 14 MHz for small signals.

At the device and materials level, we have designed, fabricated, and characterized arrays of modulators in the GaAs system operating at a wavelength of 980 nm. The modulators in such an array have simultaneously achieved, over a 2.4 mm × 2.4 mm tested area, an impressive combination of bandwidth (> 1 GHz), contrast ratio (13:1), drive voltage (9 V), insertion loss (4.5 dB), and array uniformity, as tailored to operation in our photonic neural-network system architectures. We have also demonstrated flip-chip bonding of silicon drive chips to GaAs-based modulator chips to show feasibility of hybrid integration of these devices into a smart-pixel neuron-unit array. Our unique flip-chip bonding process has been shown to provide bond strengths of more than double the highest previously published results in the same material, and collaborative efforts to transition the process to industry are underway.

We conclude that the photonic technology needed for such large-scale neural-network implementations is rapidly progressing. Smart-pixel arrays for neuron units and other functions are feasible, using the technology we describe. This technology allows for a flexibility in signal-processing function and device placement by means of VLSI design-and-fabrication processes. Arrays of modulators in the GaAs system with characteristics tailored for these kinds of photonic systems are fabricable. We project that such smart-pixel spatial-light modulator arrays can be fabricated at numbers of pixels and scale sizes sufficient for implementation of very large scale neural networks. At the systems architecture and neural-network model level, we have shown that in theory multilayer neural-network algorithms like backward error propagation can be used for learning and run to convergence on a photonic architecture that uses photorefractive holographic interconnections (under assumptions typical of photorefractive-based interconnection systems). The theory also identifies the crucial experimental parameters for ensuring such convergence. We likewise project that this technology is applicable to many other neuron-unit and neural-network models, based in part on our investigations into examples of other such models. Spin-offs into other (non-neural) applications and other technology areas can also be anticipated; examples include volume memories, interconnections, and displays.

2.2 Publications and research personnel

Publications included peer-reviewed conference and journal papers for archival dissemination of the research results, as detailed in Sect. 6. Conference and workshop talks were also given for immediate and interactive dissemination of results, as detailed in Sect. 7.1.

Personnel included three faculty principal investigators, two part-time post-doctoral research staff, a number of Ph.D. graduate students, and administrative and support staff. Details are given in Sect. 4.

3. Introduction and Overview

The purpose of this effort has been to enable dramatic improvements in the hardware performance of artificial neural-network systems by the intelligent incorporation of optical technology. We have concentrated on a long-term goal of the implementation of flexible, modular, large-scale, highly parallel, nonlinear-analog neural networks, using a combination of optics and electronics.

Our approach has been one of designing building blocks that implement functions common to many neural-network models, as well as architectures that use these building blocks in ways that can implement broad classes of neural-network models. As such, from the architecture point of view, we have focused on a single "module" that can implement generalizable neuron units and interconnections. The module is cascadable and provides for fully parallel input/output of analog signals. We have designed the architecture to have the capability for bipolar analog weights and neuron unit activations; the building blocks to be generalizable to different neural network models and learning algorithms; the system to be scalable to large neural networks with high connectivity; and the different components within the architecture to be mutually compatible. These criteria have led to the algorithm, system, component, and material research and development of this effort.

The basic photonic system we have developed and studied incorporates the following set of components and features (building blocks). First, a 2-D array of individually coherent but mutually incoherent sources illuminates the photonic system. The key features of this array that are exploited in the system are the spatial distribution of sources and their mutual incoherence. Second, a set of optoelectronic smart-pixel spatial-light modulators (SLM's) are incorporated and are all based on the same flexible technology: a silicon chip that incorporates detection and nonlinear analog electronic signal processing functions, and a gallium-arsenide-based chip that incorporates optical modulation. These optoelectronic smart-pixel SLM's can implement a parallel array of neuron units, a parallel array of error-term-calculation units, and a parallel array of training-term generators, each having (where desired) optical input and optical output. Third, an optical interconnection system is based on a volume hologram for storage of interconnection patterns and weights, implemented in a real-time photorefractive material for adaptive neural-network processing, or in a fixed thick (or stratified) material for nonadaptive (*e.g.*, pre-trained) neural networks. The double angularly multiplexed geometry provides for a number of advantages (outlined below).

The optoelectronic spatial light modulator design incorporates a number of characteristics that are key to the neural-network system. The utilization of 2 detectors and 2 modulators per pixel enables the implementation of bipolar

signals at the input to, and output from, each neuron unit (or training- or error-term pixel). The use of VLSI analog electronics within each pixel permits flexible neuron-unit functionality (*e.g.*, differencing, variable-gain sigmoid, analog storage, and derivative functions) in a small chip area (100 μm × 100 μm per pixel or neuron unit at current prototype sizes) with low power consumption. It also permits generalizability to other neuron-unit models and training algorithms. In addition, by making use of bump bonding technology, such spatial light modulators are fabricable in a hybrid package of silicon (for detectors and electronics) and GaAs-based (for modulators) materials that allows for a near-term technology insertion point.

The multiple-quantum-well optical modulators employed in the optoelectronic smart-pixel SLM's also incorporate some key unique features. The InGaAs/GaAs material system permits modulation at a wavelength of approximately 980 nm, at which the semi-insulating GaAs substrate is transparent. The optical beams can then pass through the substrate to reach the (reflection) modulators on the back of the GaAs substrate, as well as to reach the detectors on the front surface of the silicon substrate. This in turn permits direct flip-chip bonding of the silicon and GaAs chips, avoiding the need for either incorporation of relatively undeveloped via technology or full-scale monolithic integration in GaAs. It also allows an inverted Fabry-Perot cavity structure, relaxing the design constraints of the critical modulator element (and associated integrated mirrors). We have now shown that this modulator design can provide an excellent compromise among high contrast ratio, low insertion loss, and high array uniformity. In addition, at the system level, the use of an array of modulators as opposed to sources for the smart-pixel arrays is crucial for a fully adaptive system, in that mutual coherence is needed among the different pixels within some of the arrays, and within pre-defined groups of pixels that span more than one array.

The source array enables the interconnection system geometry and its resulting beneficial features. Source arrays at this wavelength of approximately 980 nm have been fabricated and demonstrated at the research level by multiple groups, including those at Bellcore and AT&T Bell Laboratories. These source arrays comprise 2-D arrays of surface emitting laser diodes. Other techniques also exist for generating an array of sources that exhibit the appropriate coherence and incoherence requirements. Many utilize a single coherent source and an optical device or system to generate an array of mutually incoherent beams. These techniques can typically operate with visible as well as infrared wavelengths, and enable laboratory demonstrations of the photonic neural network components and systems described herein.

The holographic interconnection architecture has been central to our approach, and entails some key architectural features. First, it exploits both the self-coherence and mutual-incoherence properties provided by the source array to avoid recording crosstalk terms during weight updates. Second, its

interconnection system incorporates a double angular multiplexing arrangement that eliminates fan-in beam degeneracy, reducing crosstalk during readout. Third, it provides for incoherent summation among fan-in terms at each detector in the interconnection-output plane, providing linear utilization of the detection-system dynamic range and relatively robust incoherent-system tolerances during readout. And fourth, the interconnection architecture is unique in that it permits in principle single-step copying of the entire interconnection pattern stored in a volume hologram. This is extremely useful for duplicating a neural network that has learned (perhaps at great expense in time or in hardware) a given processing or recognition function.

4. Technical Summary of Significant Work Accomplished

This section first states the objectives of the effort (Sect. 4.1), and then gives brief summaries of the significant work accomplished on key topic areas within this effort (Sect. 4.2-4.7). Much more detail can be found in the publications (listed in Sect. 6 and 7.1), and in theses referenced throughout this section. Finally, relevance and potential applications of this work are summarized (Sect. 4.8), as well as suggestions for future investigations (Sect. 4.9).

4.1 Objectives

The objectives of this effort have been to:

(1) Develop a photonic technology to implement classes of adaptive neural network models using realizable hardware;
(2) Realize and exploit high temporal bandwidths and spatial parallelism; and
(3) Assess the fundamental and technological potential and limitations of the photonic hardware.

These objectives have been stated in presentations of our effort at the ARPA Artificial Neural Networks Technology Program Reviews throughout the grant period, and have not changed.

The approach and rationale for the first objective is discussed above in Sect. 3, and many of the results below and in the publications answer to this objective directly.

The second objective deals with spatial parallelism and temporal bandwidths. A high degree of spatial parallelism is achievable by direct extension of the approach introduced above; for example, 100 μm × 100 μm neuron-units allow 9×10^4 neuron units in a 3 cm × 3 cm chip; by scaling down feature sizes, even higher numbers should be possible. Issues in this scaling-up process include uniformity of response (addressed below), storage capacity of the holographic medium (addressed in other works), and scalability of the overall algorithm and hardware effects such as noise, optical throughput efficiency, etc. (addressed in publications referenced in Sect. 4.2, below, for the algorithm, overall system, and SLM detection; addressed in work under the previous grant for the optical interconnection system by itself). High temporal bandwidths are possible given the SLM technology we are using, and experimental results indicating the potential thereof are summarized in the sections below. Potential uses of these high temporal bandwidths include high-speed nonadaptive neural-network computing; optimization of exposure energy and time during learning

in an adaptive neural-network system; and incorporation of neural network models that use temporal pulse trains or binding mechanisms such as the dynamic link architecture of von der Malsburg.

Work on the third objective has been continual throughout our research, and is included in essentially all of the subtopics we have investigated. Results are described in some of the summaries of results below, as well as in the publications listed below.

4.2 Photonic multilayer backward-error-propagation learning: theory and performance analysis

Work on the topic of this subsection was funded in part by the grant being reported on, and in part by the related AASERT grant, No. F49620-93-1-0445. The AASERT grant funded the key graduate student that worked on this topic.

We have chosen a specific class of neural-network models and learning algorithms to analyze in detail, in order to assess the capability and limitations of the photonic neural-network hardware toolbox and architecture. The model class chosen was that of multilayer feedforward neural networks, using bipolar-input, unipolar-output analog inner-product neuron units (using a sigmoidal response function), trained with backward error propagation. Reasons for this choice are that the learning algorithm is known to be non-trivial to implement in analog hardware, and is representative of the broad class of outer-product learning rules and therefore maps into holographic-interconnection systems; the neural-network model requires cascadability of stages; and the model represents a powerful neural network trained by a commonly used (albeit nonideal) training algorithm.

In this subsection, we summarize some of the key results we have obtained in a theoretical investigation of the learning properties of such a neural-network model implemented on two broad classes of photonic hardware architectures. Additional detail can be found in the publications.

- We have analytically determined that the backward-error-propagation learning algorithm, implemented on either of two broad classes of photonic multilayer feedforward neural networks, has a well defined region of convergence. This region of convergence can be represented in a neural learning-parameter space (in terms of neural-network signals and quantities), *i.e.*, as a function of a learning-rate parameter and a weight-decay parameter. We consider this region of convergence for two broad classes of photorefractive-based optical neural network architectures. The first class uses electric field amplitude encoding of signals and weights in a fully coherent system, whereas the second class uses intensity encoding of signals and weights in an incoherent/coherent system. Under typical assumptions

on the grating formation in photorefractive materials used in adaptive optical interconnections, we have derived a set of conditions that are sufficient for such a network to operate within the region of convergence. The results are verified empirically by simulations of the XOR sample problem.

- We have expressed the above convergence conditions in terms of physical (as opposed to the above neural-network) quantities. We have shown that the neural-space learning parameters are directly related to two important physical design parameters: system gain and exposure energy. The system gain determines the ratio of the learning-rate parameter to decay-rate parameter, and the exposure energy determines the size of the decay-rate parameter. We have concluded that convergence can be guaranteed (assuming no spurious local minima in the error function) by using a sufficiently high gain and a sufficiently low exposure energy.

- We have also investigated the effects of noise, due to the optical signal itself as well as due to the spatial-light-modulator detection system. We have shown that in terms of minimizing the neural-network system noise, there is an optimum ratio of photorefractive-readout exposure energy (used during forward and backward neural-network passes) to photorefractive-recording exposure energy (used during weight updates), under the assumption that these two exposure energies are constant throughout neural-network learning.

- We have shown that the system noise places additional constraints on the learning convergence, so for the network learning process to converge, it must both satisfy the above convergence condition, and have noise variances that are sufficiently small. Based on these conditions, we extrapolated the simulation results of the XOR sample problem to estimate the maximum size network that can be implemented in a 1.5 cm $\times$ 1.5 cm $Fe:LiNbO_3$ photorefractive crystal. For the class of architecture that uses electric field amplitude encoding of signals and weights, and for problem types in which the convergence condition scales inversely with network size, the maximum size network that can be implemented is on the order of the crystal's usable storage capacity.

- The effects of modulator contrast ratio on system convergence were studied. Simulation results indicate how the sufficient condition for convergence depends on contrast ratio for both classes of architectures. For the class of architecture that uses field amplitude encoding of signals, we have also derived this dependence analytically. For the architecture that uses intensity encoding of signals, our simulation results indicate that the sufficient condition for convergence is relatively weakly dependent on the modulator contrast ratio. These results, in conjunction with the experimental contrast ratios achieved with our modulator arrays (Sect. 4.4), indicate that this

modulator array technology is feasible for use in these photonic architectures for implementing certain classes of neural networks.

More detail on the above results can be found in [Publication Nos. 9, 16] and in the following thesis:

>> G. C. Petrisor, *Convergence of Backward-Error-Propagation Learning in Photorefractive Crystals*, Ph. D. Thesis, USC-SIPI Report No. 303 (Signal and Image Processing Institute, University of Southern California, Los Angeles, California 90089-2564, December 1996). Copies are available for purchase by contacting the Signal and Image Processing Institute, Phone: (213) 740-4145, Fax: (213) 740-4651, e-mail: gloria@sipi.usc.edu.

Significance to field and relationship to original goals. It was previously unknown whether neural-network learning would work in general on photorefractive-based systems. This result provides a general framework within which to understand how to build and set up a photorefractive-based photonic system that will correctly implement multilayer neural-network learning. Issues such as spatial light modulator gain requirements, acceptable levels of photorefractive decay and detector noise, and optimal exposure levels can now be evaluated for a given application. This clearly helps delimit the fundamental potential and limitations of the hardware, as well as helps ascertain the bandwidths that are preferable for a given application, and guide development work on the hardware components (*e.g.*, in order to exceed spatial light modulator gain and laser output power requirements).

4.3 Optical radial basis function neural networks: analysis and development

Work on the topic of this subsection was funded in part by the grant being reported on, and in part by the related AASERT grant, No. F49620-93-1-0445. The AASERT grant funded the key graduate student that worked on this topic.

In order to address the issue of flexibility and extendibility to other neural network models, we considered the question of how well this same toolbox of photonic components and interconnections could be applied to other models. One such model we investigated was a 2-layer radial-basis-function neural network that is designed to be capable of implementing nonparametric density-function estimation, pattern recognition based on density estimation, and fuzzy logic processing.

Key results we have achieved on this topic are summarized in the following. Additional detail can be found in the reference listed below.

- We have developed and analyzed an optical system that implements a radial-basis-function neural network, based on pre-recorded fixed holograms, for pattern recognition applications. The patterns to be recognized are recorded onto planar or volume holograms; the holograms are then fixed and used in the radial-basis-function neural network for high-speed parallel pattern-recognition tasks.

- We have developed a continuous version of the well-known k-nearest neighbors density-function estimation and pattern-recognition algorithm (which we call *continuous k-nearest neighbors*, or *C-kNN*), in order to tailor it to implementation on parallel analog optical or electronic hardware. Use of the above optical system for this purpose has been analyzed. Simulations of this C-kNN algorithm have yielded positive results, in that its performance is comparable to the conventional k-nearest-neighbors algorithm.

Significance to field and relationship to original goals. For applications in which all learning is done ahead of time, this new architecture provides a direct way of recording input training patterns into holograms and then using them in the pattern recognition system. Recognition is fast and done in parallel. The system can be efficiently implemented on analog hardware. This architecture adds another class of neural network models that can be implemented with basically the same underlying set of photonic hardware components that are described elsewhere in this report.

More detail on the above results can be found in the following thesis:

>> A. A. Goldstein, *Scalable Photonic Neural Networks for Real-Time Pattern Classification*, Ph. D. Thesis, USC-SIPI Report (Signal and Image Processing Institute, University of Southern California, Los Angeles, California 90089-2564, February 1997). Copies are available for purchase by contacting the Signal and Image Processing Institute, Phone: (213) 740-4145, Fax: (213) 740-4651, e-mail: gloria@sipi.usc.edu.

4.4 Modulator arrays for neuron-unit outputs: fabrication and characterization

The modulator arrays provide optical parallel outputs for arrays of optoelectronic neuron units, error-term calculation units, or training-term-generation units.

Our approach has been the use of multiple-quantum-well modulators in an asymmetric Fabry-Perot cavity, implemented in a strained layer system operating at 980 nm wavelength, with readout through the semi-insulating GaAs substrate. Key issues have included the fabrication of arrays of modulators that simultaneously exhibit a sufficient degree of the following competing

requirements: high contrast ratio, high uniformity across the array, low insertion loss, and low drive voltage. Temporal bandwidth is also of interest for the potential higher bandwidth applications mentioned above in Sect. 4.1.

Key results we have achieved on this topic are summarized in the following. Additional detail can be found in the publications.

- We have achieved high contrast modulation and 980 nm operation in the highly uniform (5×5 mm^2) central region of the molecular-beam-epitaxy (MBE) grown wafer. This allows the fabrication of large 2-D arrays (100×100 on a 50 μm center-to-center pitch) of light modulators. The modulator design follows that reported for the previous period. It is based upon use of 99% back mirror reflectance as provided by dielectric stack Bragg mirrors and comprises 35 $Al_{0.26}Ga_{0.74}As/In_{0.20}Ga_{0.80}As$ quantum wells and a 6.5 pair AlAs/GaAs (~70% reflectance) front mirror. For testing and growth characterization purposes, the back mirror material has been a metal and was changed from gold (94% reflectance) to silver (97% reflectance) to allow close comparison between test structures and devices employing dielectric stack back mirrors. A detailed uniformity study was made over a 2.4×2.4 mm^2 area within the 5×5 mm^2 highly uniform region. This area was selected to be a bit larger than the 1.6×1.6 mm^2 active area of the latest Si neuron driver chips. Over this area (using Ag back mirrors) the contrast ratio was measured to be 13:1 ± 4:1, the contrast ratio optical bandwidth was 3.2 nm FWHM, the dynamic range 35 ± 3% (insertion loss 4.5 dB), the operating voltage 9.0 V, and the operating wavelength 980.7 ± 0.3 nm. These values meet or exceed specifications and are state-of-the-art for III/V semiconductor light modulators operating at 980 nm. The insertion loss, however, is higher than desired. This is due to a combination of absorption losses in the 350 μm thick semi-insulating GaAs substrate, the use of metallic (97%) rather than dielectric (99%) high reflectance mirrors and absorption losses in the multiple quantum well cavity itself.

- Modulators sized to match the silicon driver-chip array dimensions, *i.e.*, 30 μm × 40 μm modulators on 50 μm center-to-center pitch have been fabricated without complication. Individual characterization of such small pixels without damaging the (20×30 μm^2) back mirror/electrical contact with test probes is difficult. The devices successfully tested show the same performance characteristics as larger test pixels. Larger size pixels (150 μm center-to-center) were used for the detailed uniformity study described above. Devices with smaller pixel sizes are used for hybrid integration with the 16×16 2-D neuron-unit array chip.

- We have demonstrated modulation bandwidths in excess of 1 GHz, drive current limited.

Significance to field and relationship to original goals. Although many strides have been made in increasing the contrast ratio of individual multiple-quantum-well modulator elements, making an array that is reasonably uniform without unduly compromising the contrast ratio is crucial to photonic neural network systems and had previously not been achieved. In addition, for compatibility with silicon driver chips, the drive voltages need to be down to the range of 10 volts or less. These new results show that such uniform arrays are feasible, can have reasonable contrast ratios for analog neural network systems, and can be driven with silicon-compatible drive voltages. The high bandwidths also allow temporal-domain neural network models to be considered, such as von der Malsburg's dynamic link architecture.

4.5 Silicon driver chips for neuron-unit signal processing and inputs: fabrication and characterization

The silicon driver chips provide optical input, electronic receiver, electronic nonlinear-analog signal processing, and modulator-driving functions. Our approach has been to use computer techniques to design analog photonic and electronic circuitry in a smart-pixel-array layout geometry, fabricated using standard silicon-VLSI processes. These steps allow us to achieve a very nonstandard implementation of amplification, differencing, and signal-shaping functions in a small chip area with low power dissipation, yet sufficiently high bandwidth.

Key results we have achieved on this topic are summarized in the following. Additional detail can be found in the publications.

- We have fabricated and successfully demonstrated 16×16 arrays of photodetectors, control circuitry, and modulator drivers in silicon. The control circuitry includes a dual-rail input, dual-rail output sigmoidal neuron-unit activation function. This latest CMOS design also has guard rails for the photodetectors to improve the signal-to-noise ratio and reduce crosstalk between adjacent pixels. The Si neuron-unit-array chips are undergoing electrical characterization and are designed for small signal bandwidths of 14 MHz. The area for each neuron unit, including input detectors and output pads, is $100\ \mu m \times 100\ \mu m$, using the $1.2\ \mu m$ N-well CMOS process.

Significance to field and relationship to original goals. These results demonstrate that the input, subtraction, sigmoid, and output functions are feasible using analog circuitry in an area appropriate for arrays of large numbers of neuron units in a smart-pixel spatial light modulator.

4.6 Flip-chip bonding for hybrid integration of neuron-unit arrays: demonstration and characterization

The flip-chip bonding process is key to the fabrication of complete optoelectronic smart-pixel spatial light modulators (SLM's), and allows implementation of full neuron-unit arrays (or arrays of pixels that implement other functions needed in adaptive or nonadaptive neural networks). Our SLM geometry uses one bump bond for each output modulator (thus two bonds per pixel in our bipolar neuron-unit array), allowing fully parallel communication from the Si chip to the GaAs chip. We have developed an indium bump process that uses Velcro-like adhesion for extremely high bond strength. Key results we have achieved on this topic are summarized in the following.

- We have flip-chip bonded silicon driver chips to GaAs modulator chips, by evaporating an 8×8 array of $30\ \mu m \times 30\ \mu m \times 5\ \mu m$ high indium bumps onto aluminum contacts and using a thick-film multilayer photolithographic process. The resulting bonds each have a contact resistance less than a few ohms. A 16×16 array using $10\ \mu m \times 10\ \mu m \times 7\ \mu m$ high indium bumps has been demonstrated.

- We have flip-chip bonded directly to the modulator elements on the GaAs chip. As desired, modulator operation was not affected by the bonding process.

- Mechanical tests of the flip-chip bonded device show the bonds to be highly stable and secure.

The USC Velcro-like indium bump process uses thermal evaporation to achieve typical evaporation rates of 25 to 30 Å per second. For the chip sets referred to in this project the nominal indium contact bump height was $\sim 12\ \mu m$, with surface peak-to-valley variations of up to $6\ \mu m$.

- The ultimate tensile strength for these bump contacts has been measured to be 139.6 kg/cm^2. This value was obtained from tensile pull tests resulting in 2.08 kg tensile yield force for 304 bumps, each with a $70\ \mu m \times 70\ \mu m$ cross section. Two comparisons can put this result into perspective. The theoretical bulk indium yield strength is 16.32 kg/cm^2, more than a factor of eight smaller than our bond strength. Compared with other indium contact bonding techniques, our novel indium contact process for flip-chip applications is nearly two and half times stronger than any other competing indium contact process reported in the literature.

Significance to field and relationship to original goals. Hybrid integration of photonic Si and GaAs chips promises to allow the electrical integration densities and functionalities of silicon VLSI in a smart-pixel spatial light

modulator device. Demonstrating the basic principles of this bonding opens the door for the design and fabrication of arrays of neuron units operating at high bandwidth, with neuron unit functionality designable using standard Si computer-aided design and layout techniques, and Si chips fabricated with standard Si fabrication process steps.

4.7 Compact optical system for the photonic neural network: design

In order to evaluate the potential for physically small implementations of the adaptive neural-network architecture, we have taken the first step in the process of designing a compact version of the photonic neural-network architecture.

- We have performed a phase-one design and analysis of a compact optical system for implementation of adaptive, volume-holographically interconnected, photonic neural networks.

Significance to field and relationship to original goals. For the hardware to be realizable, it will ideally be small in size and robust to physical movement and jarring. The first-phase design and analysis is a significant step toward these goals. Additionally, we point out that this is not a requirement for all applications. For the case of training a master network and then making copies of the interconnection for use in nonadaptive systems, only the nonadaptive system need be physically small and robust. This is a much simpler requirement to meet because the incoherent nature of the readout process reduces the required stability of the system, and the lack of a need for a holographic recording geometry further reduces the complexity of the system.

4.8 Relevance and potential applications

This work relates directly to the realization of physically compact, high-performance adaptive and nonadaptive neural networks. Such neural networks have potential application in pattern recognition including automatic target recognition, to intelligent assistants (*e.g.*, pilot's assistant and in-the-field troop assistants), and to autonomous control of mobile craft and weaponry.

Additional potential spin-off technologies include high-speed, parallel input and output devices for optical volume memory; high-capacity volume holographic storage devices; built-in intelligent renderers and drivers for graphical displays; high-density optical interconnections; and true three-dimensional real-time computer displays.

4.9 Suggestions for future investigation

Additional issues that are worthy of future investigation include, first, wavelength compatibility of the source array, SLM, and volume holographic medium. Work has been progressing on this front in the areas of infrared-sensitive photorefractive materials on the one hand, and visible-wavelength laser-diode sources on the other. Second, the true scaling properties of the photonic neural-network system as a whole cannot yet be predicted. Our understanding of the operation of such a full system is now much more complete, *i.e.*, includes multilayer effects and salient characteristics of the different components when used in a nonlinear system that has inherent feedback due to the learning and computing processes. The next logical step (which so far has been intractable) is to analyze the learning-convergence and performance properties when the number of neuron units and interconnections grows to very large numbers. Finally, the scenario of large-scale pre-trained networks that operate with little or no adaptation is worthy of additional investigation. Such networks can provide a nearer-term technology insertion with a broad range of potential applications.

5. Personnel Supported

The following personnel were financially supported in part by this grant, in compensation for work performed on this effort. Almost all of these personnel were part time on this effort.

5.1 Faculty

B. Keith Jenkins, Anupam Madhukar, and Armand R. Tanguay, Jr. were supported as principal and co-principal investigators. Ping Chen was supported as a research investigator. Christoph von der Malsburg was also associated with this effort through technical collaborations, although he was not financially supported by this grant.

5.2 Administrative staff

Stacy Bant, Gloria Halfacre, Elaine Stearns, Karen Tierney, and Linda Varilla were supported as administrative support staff. Mariela Ortiz and Cecilia Crawford were supported as student support staff.

5.3 Post-doctoral research staff

Chris Kyriakakis and John Rilum were supported as post-doctoral level investigators.

5.4 Graduate students

Kartik Ananthanarayanan, Robert Cartland, Jun Chen, Wei Chen, Scott DeMars, Edward Herbulock, Ke-Zhong Hu, Chingchu Huang, Jong-Je Jung, Nobuhiko Kobayaashi, Kuang-Yu Li, Patrick Nasiatka. Sabino Piazzolla, and Nainjeet Ramlagan were supported as Ph.D. graduate students.

6. Publications

Below are listed peer-reviewed publications that resulted at least in part from the work on this grant.

1. G. P. Nordin, P. Asthana, A. R. Tanguay, Jr., and B. K. Jenkins, ``Analysis of Weighted Fan-out/Fan-in Volume Holographic Interconnections," Topical Meeting on Diffractive Optics: Design, Fabrication, and Applications, *1992 Technical Digest Series*, Vol. 9 (Optical Society of America, Washington, D.C., 1992), pp. 165-167.

2. B. K. Jenkins, A. R. Tanguay Jr., and A. Madhukar, "Photonic Technology for Densely-Interconnected Neural Networks," *1992 Digest of Papers, Government Microcircuit Applications Conference*, Las Vegas, Nevada, November 9-12, 1992, pp. 563-566 (Defense Technical Information Center, Washington, D.C., 1992).

3. K. Hu, R. Cartland, Li Chen, K. Kaviani, P. Chen and A. Madhukar, "Ex-situ cavity phase tuning of InGaAs/AlGaAs multiple quantum well based on inverted asymmetric Fabry-Perot reflection modulators," *Spatial Light Modulators and Applications Topical Meeting*, 15-17 March 1993, Palm Springs, CA; 1993 Technical Digest Series, Vol. 6 (Optical Society of America, Washington, D.C., 1993), pp. 170-173.

4. P. Asthana, G. P. Nordin, A. R. Tanguay, Jr., and B. K. Jenkins, "Analysis of Weighted Fan-out/Fan-in Volume Holographic Optical Interconnections," *Applied Optics*, Vol. 32, No. 8, pp. 1441-1469 (10 March 1993).

5. A. R. Tanguay, Jr., A. Madhukar and B. K. Jenkins, "Hybrid Silicon/Gallium Arsenide Inverted Fabry-Perot Cavity MQW Spatial Light Modulators," invited paper, *Spatial Light Modulators and Applications Topical Meeting*, 15-17 March 1993, Palm Springs, CA; 1993 Technical Digest Series, Vol. 6 (Optical Society of America, Washington, DC, 1993) p. 127.

6. G. C. Petrisor, S. Piazzolla, G. P. Nordin, B. K. Jenkins, A. R. Tanguay, Jr., "Volume Holographic Interconnection and Copying Architectures Based on Incoherent/Coherent Source Arrays," *Proc. Fourth International Conference on Holographic Components, Systems and Applications*, Neuchatel, Switzerland, 13-15 September 1993, IEE Conference Publication No. 379 (Institution of Electrical Engineering, London, 1993) pp. 21-26.

7. Z. Karim, C. Kyriakakis, A. R. Tanguay, Jr., K. Hu, L. Chen, and A. Madhukar, "Externally-deposited phase-compensating dielectric mirrors for

asymmetric Fabry-Perot cavity tuning," *Applied Physics Letters*, Vol. 64, 2913 (1994).

8. A. A. Goldstein, G. C. Petrisor, and B. K. Jenkins, "Gain and exposure scheduling to compensate for photorefractive neural-network weight decay," submitted October 3, 1994, accepted December 1994, *Optics Letters*, Vol. 20, No. 6, pp. 611-613 (March 15, 1995).

9. G. C. Petrisor, A. A. Goldstein, E. J. Herbulock, B. K. Jenkins, and A. R. Tanguay, Jr., "Convergence of backward error propagation learning in photorefractive crystals," in *Optical Computing*, 1995 OSA Technical Digest Series, Vol. 10, pp. 158-160, March 1995 (Optical Society of America, Washington DC, 1995).

10. C. Huang, B. K. Jenkins, and C. B. Kuznia, "Weighted space-variant local interconnections based on micro-optic components: crosstalk analysis and reduction," in *Optical Computing*, 1995 OSA Technical Digest Series, Vol. 10, pp. 280-282, March 1995 (Optical Society of America, Washington DC, 1995).

11. C. Kyriakakis, Z. Karim, A. R. Tanguay, Jr., R. F. Cartland, A. Madhukar, S. Piazzolla, B. K. Jenkins, C. B. Kuznia, A. A. Sawchuk, and C. v. d. Malsburg, "Photonic implementations of neural networks," in *Optical Computing*, 1995 OSA Technical Digest Series, Vol. 10, pp. 128-130, March 1995 (Optical Society of America, Washington DC, 1995).

12. B. Keith Jenkins and A. R. Tanguay, Jr., "Optical Architectures for Neural Network Implementations," in M. A. Arbib, Ed., *The Handbook of Brain Theory and Neural Networks*, pp. 673-677 (MIT Press, Cambridge, Mass., 1995).

13. A. R. Tanguay, Jr. and B. K. Jenkins, "Optical Components for Neural Network Implementations," in M. A. Arbib, Ed., *The Handbook of Brain Theory and Neural Networks*, pp. 677-682 (MIT Press, Cambridge, Mass., 1995).

14. Z. Karim, C. Kyriakakis, A. R. Tanguay, Jr., R. F. Cartland, K. Hu, L. Chen and A. Madhukar, "Post-growth tuning of inverted cavity InGaAs/AlGaAs spatial light modulators using phase compensating dielectric mirrors", *Applied Physics Letters*, Vol. 66, No. 21, (1995).

15. A. A. Goldstein, G. C. Petrisor, and B. K. Jenkins, "Gain and exposure scheduling to compensate for photorefractive neural-network weight decay," *Optics Letters*, Vol. 20, No. 6, pp. 611-613 (15 March 1995).

16. G. C. Petrisor, A. A. Goldstein, B. K. Jenkins, E. J. Herbulock, and A. R. Tanguay, Jr., "Convergence of backward-error-propagation learning in photorefractive crystals," *Applied Optics*, Vol. 35, No. 8, pp. 1328-1343 (10 March 1996).

7. Interactions/Transitions

7.1 Participation/presentations at meetings, conferences, seminars

1. G. P. Nordin, P. Asthana, A. R. Tanguay, Jr., and B. K. Jenkins, "Analysis of Weighted Fan-out/Fan-in Volume Holographic Interconnections," Topical Meeting on Diffractive Optics: Design, Fabrication, and Applications, (1992) (associated with written paper No. 1 above).

2. B. K. Jenkins, A. R. Tanguay Jr., and A. Madhukar, "Photonic Technology for Densely-Interconnected Neural Networks," *Government Microcircuit Applications Conference*, Nevada, November 9-12, (1992) (associated with written paper No. 2 above).

3. K. Hu, R. Cartland, Li Chen, K. Kaviani, P. Chen and A. Madhukar, "Ex-situ cavity phase tuning of InGaAs/AlGaAs multiple quantum well based inverted asymmetric Fabry-Perot reflection modulators," *Spatial Light Modulators and Applications Topical Meeting*, 15-17 March (1993), Palm Springs, CA, 1993, paper SWB2 (associated with written paper No. 3 above).

4. A. R. Tanguay, Jr., A. Madhukar and B. K. Jenkins, "Hybrid Silicon/Gallium Arsenide Inverted Fabry-Perot Cavity MQW Spatial Light Modulators," invited paper, *Spatial Light Modulators and Applications Topical Meeting*, 15-17 March, (1993), Palm Springs, CA, paper STuB3 (associated with written paper No. 5, above).

5. B. K. Jenkins A. Madhukar, A. R. Tanguay, Jr., L. Chen, K. Z. Hu, Z. Karim, C. Kyriakakis, G. P. Nordin, G. C. Petrisor, S. Piazzolla and D. Su, "Photonic Neural Network Implementations Based on Incoherent/Coherent Holographic Interconnections," *Symposium on Optoelectronic Neural Networks*, SPIE, 4-8 July, (1993), San Diego, CA, paper No. 2026-40.

6. Jun Chen, R. Cartland, K. Hu, L. Chen, P. Chen and A. Madhukar, "Post-growth Cavity Tuned Strained InGaAs/AlGaAs MQW based Inverted Asymmetric Fabry-Perot Reflection Spatial Light Modulators", *North American Conference on Molecular Beam Epitaxy,* Stanford, CA, Sept. 13-15, (1993).

7. G. C. Petrisor, S. Piazzolla, G. P. Nordin, B. K. Jenkins, A. R. Tanguay, Jr., "Volume Holographic Interconnection and Copying Architectures Based on Incoherent/Coherent Source Arrays," *Fourth International Conference on Holographic Components, Systems and Applications*, Neuchatel, Switzerland, 13-15 September, (1993) (associated with written paper No. 6, above).

8. B. K. Jenkins, A. Madhukar, and A. R. Tanguay, Jr., "Photonic technology development for densely interconnected neural networks," *ARPA Artificial Neural Network Technology Program Review*, Key West, Florida, December 6-8, (1994).

9. G. C. Petrisor, A. A. Goldstein, E. J. Herbulock, B. K. Jenkins, and A. R. Tanguay, Jr., "Convergence of backward error propagation learning in photorefractive crystals," presented at the *Topical Meeting on Optical Computing*, Optical Society of America, Salt Lake City, Utah, March 13-16, (199)5 (associated with written paper 9, above).

10. C. Huang, B. K. Jenkins, and C. B. Kuznia, "Weighted space-variant local interconnections based on micro-optic components: crosstalk analysis and reduction," presented at the *Topical Meeting on Optical Computing*, Optical Society of America, Salt Lake City, Utah, March 13-16, (1995) (associated with written paper 10, above).

11. C. Kyriakakis, Z. Karim, A. R. Tanguay, Jr., R. F. Cartland, A. Madhukar, S. Piazzolla, B. K. Jenkins, C. B. Kuznia, A. A. Sawchuk, and C. v. d. Malsburg, "Photonic Implementations of neural networks," presented at the *Topical Meeting on Optical Computing*, Optical Society of America, Salt Lake City, Utah, March 13-16, (1995) (associated with written paper 11, above).

12. J. J. Jung, G. P. Nordin, A. R. Tanguay, Jr., "Effect of buffer layer thickness variations on stratified volume holographic optical elements," *Annual Meeting of the Optical Society of America*, Dallas, Texas, Oct. 2-7, (1994), paper MDD2.

13. K. Ananthanarayanan, C. H. Chen, S. DeMars, A. A. Goldstein, C. C. Huang, D. Su, C. B. Kuznia, C. Kyriakakis, Z. Karim, B. K. Jenkins, A. A. Sawchuk, A. R. Tanguay, Jr., "Multilayer Electronic/Photonic Multichip Modules with Vertical Optical Interconnections", *Annual Meeting of the Optical Society of America*, Portland, Oregon, (1995), paper ThCC7.

14. C. B. Kuznia, C. C. Huang, K. Ananthanarayanan, C. H. Chen, A. A. Sawchuk, "Micro Diffractive Optical Elements for Smart Pixel Fanout Interconnections", *Annual Meeting of the Optical Society of America*, Portland, Oregon, (1995).

15. R. F. Cartland, P. Chen and A. Madhukar, "High Contrast, large area, inverted, 2D spatial light modulators for flip-chip bonding to silicon microelectronic driver chips", paper TuCC3, *Annual Meeting of the Optical Society of America* , Rochester, NY, (1996).

16. R. F. Cartland and A. Madhukar, "High Contrast, 2D Spatial Light Modulators (SLMs) Using InGaAs/AlGaAs Quantum Wells Operating at 980

nm", paper SMD4, *Spring Topical Meeting on Spatial Light Modulators, Optical Society of America,* Lake Tahoe, NV, March 17-19, (1997).

7.2 Transitions

1. Talks are currently underway for licensing the original patent on the holographic techniques employed in this effort, "Incoherent/coherent multiplexed holographic recording for photonic interconnections and holographic optical elements," USPTO Patent No. 5,121,231 (invented by two of the principal investigators of this grant, under partial support of the previous grant, "Photonic Technology for Implementation of Generalizable Neural Networks: A Synthetic Approach", Grant No. AFOSR-89-0466), to a small company (name withheld to maintain confidentiality).

2. We have developed a Velcro-like indium bump bonding process (described above in Sect. 4.6), and have performed this process on the flip-chip bonding of multichip modules for TRW, Inc., as part of the transitioning of this new process to industry.

8. New Discoveries, Inventions, or Patent Disclosures

8.1 Patents awarded

The following patent applications were issued by USPTO in this reporting period. The research work of this grant is based largely on them, and further develops the apparatuses described in them. They are based on inventions that were originally conceived during or before the time period of the previous grant to this one ("Photonic Technology for Implementation of Generalizable Neural Networks: A Synthetic Approach", grant No. AFOSR-89-0466).

1. B. K. Jenkins and A. R. Tanguay, Jr., "Spatial Light Modulators for Incoherent/Coherent Multiplexed Holographic Recording and Readout," assigned to University of Southern California, U. S. Pat. No. 5,285,308, file date 13 Feb. 1992 (divisional of U. S. Pat. No. 5,121,231, filed 6 Apr. 1990), issued 8 Feb. 1994.

2. B. K. Jenkins and A. R. Tanguay, Jr., "Incoherent/Coherent Source Array for Multiplexed Holographic Recording and Readout," assigned to University of Southern California, U. S. Patent No. 5,339,177, file date 13 Feb. 1992 (divisional of U. S. Pat. No. 5,121,231, filed 6 Apr. 1990), issued 16 Aug. 1994.

3. B. K. Jenkins and A. R. Tanguay, Jr., "Incoherent/Coherent Readout of Double Angularly Multiplexed Volume Holographic Optical Elements," assigned to University of Southern California, U. S. Patent No. 5,416,616, file date 8 June 1992 (continuation-in-part of U. S. Pat. No. 5,121,231, filed 6 Apr. 1990), issued 16 May 1995.

8.2 Inventions conceived, new invention disclosures, and new patent disclosures

The following divisional patent application was filed with USPTO in this reporting period; it is based on inventions conceived during the time period of the previous grant to this one ("Photonic Technology for Implementation of Generalizable Neural Networks: A Synthetic Approach", Grant No. AFOSR-89-0466).

1. B. K. Jenkins and A. R. Tanguay, Jr., "Incoherent/Coherent Double Angularly Multiplexed Volume Holographic Optical Elements," assigned to University of Southern California, filed 15 May 1995 with U. S. Patent and Trademark Office (divisional of U. S. Pat. No. 5,416,616, filed 8 June 1992, in

turn a continuation-in-part of U. S. Pat. No. 5,121,231, filed 6 Apr. 1990), pending.